Sporulation capability and amylosome conservation among diverse human colonic and rumen isolates of the keystone starch-degrader *Ruminococcus bromii*

Indrani Mukhopadhya[1], Sarah Morais[2], Jenny Laverde-Gomez[1],Paul O Sheridan[1], Alan W Walker[1], William Kelly[3], Athol V Klieve[4,6], Diane Ouwerkerk[5,6], Sylvia H Duncan[1], Petra Louis[1], Nicole Koropatkin[7], Darrell Cockburn[7], Ryan Kibler[7], Philip J Cooper[8], Carlos Sandoval[8], Emmanuelle Crost[9], Nathalie Juge[9], Edward A Bayer[2], Harry J Flint[1*]

Affiliations:

1. Microbiology Group, The Rowett Institute, University of Aberdeen, Aberdeen, UK.
2. Department of Biomolecular Sciences, The Weizmann Institute of Science, Rehovot, Israel.
3. AgResearch Limited, Grasslands Research Centre, Palmerston North, 4442, New Zealand.
4. School of Agriculture and Food Sciences, The University of Queensland, St Lucia, QLD Australia.
5. Agri-Science Queensland, Department of Agriculture and Fisheries, Brisbane, QLD Australia.
6. Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD Australia.
7. Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan, USA.
8. Hospital Cantonal "Padre Alberto Buffoni", Avenida 3 de Julio y Victor Villegas, Quininde, Esmeraldas Province, Ecuador.
9. The Gut Health and Food Safety Institute Strategic Programme, Institute of Food Research, Norwich, United Kingdom.

Running title: Comparative genomics of *Ruminococcus bromii*

***Corresponding author:** Harry J Flint - Microbiology Group, Gut Health theme, The Rowett Institute, University of Aberdeen, Aberdeen, UK.

Email: h.flint@abdn.ac.uk

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.14000

Originality-Significance Statement.

We present here a unique analysis of human colonic and rumen *Ruminococcus bromii* strains from four continents that combines genome sequencing, cultivation and molecular biology. Strikingly, the organization of degradative enzymes into 'amylosome' complexes, a feature so far found only in this 'keystone' starch-degrading species, is common to all strains and shows an extraordinary degree of conservation. This highlights the functional significance of enzyme organization in determining polysaccharide-degrading activity. We also show for the first time that this anaerobic gut symbiont is able to form oxygen-resistant spores, an ability that must be highly significant in the ability of this nutritionally specialised strict anaerobe to transmit between hosts.

Summary

Ruminococcus bromii is a dominant member of the human colonic microbiota that plays a 'keystone' role in degrading dietary resistant starch. Recent evidence from one strain has uncovered a unique cell surface 'amylosome' complex that organizes starch-degrading enzymes. New genome analysis presented here reveals further features of this complex and shows remarkable conservation of amylosome components between human colonic strains from three different continents and a *R. bromii* strain from the rumen of Australian cattle. These *R. bromii* strains encode a narrow spectrum of carbohydrate active enzymes (CAZymes) that reflect extreme specialization in starch utilization. Starch hydrolysis products are taken up mainly as oligosaccharides, with only one strain able to grow on glucose. The human strains, but not the rumen strain, also possess transporters that allow growth on galactose and fructose. *R. bromii* strains possess a full complement of sporulation and spore germination genes and we demonstrate the ability to form spores that survive exposure to air. Spore formation is likely to be a critical factor in the ecology of this nutritionally highly

specialised bacterium, which was previously regarded as 'non-sporing', helping to explain its widespread occurrence in the gut microbiota through the ability to transmit between hosts.

Introduction

Ruminococcaceae are an important family of Firmicutes bacteria within gut microbial communities (La Reau et al., 2016). These bacteria account for around 20% of the healthy human colonic microbiota based on molecular surveys (Suau et al., 2001) with two species, *Faecalibacterium prausnitzii* and *Ruminococcus bromii*, among the four most abundant contributors to the human faecal metagenome in European adults (Zhernakova et al., 2016). While *F. prausnitzii* in particular has been implicated in health maintenance (Sokol et al., 2008) a scarcity of cultured isolates, partly due to fastidious growth requirements (Herbeck and Bryant, 1974; Ze et al., 2012) has limited the information available on other Ruminococcaceae. *R. bromii* is thought to be a specialist starch-utilizing bacterium, with at least one strain showing superior ability to degrade insoluble starches when compared with other amylolytic human gut bacteria (Ze et al., 2012). Resistant starch (RS), i.e. dietary starch that escapes digestion by host amylases, often provides the largest single source of energy for microbial growth in the human colon and its fermentation is considered to provide health benefits (Nugent, 2005). Significantly, the relative abundance of *R. bromii* 16S rRNA gene sequences has been found to increase rapidly and dramatically in faecal samples from human volunteers after switching onto diets high in RS (Walker et al., 2011; Salonen et al., 2014), indicating an outstanding ability to compete for RS *in vivo*. Similar rapid population increases have been reported for *R. bromii* within the rumen microbiota in cattle fed starch-enriched diets (Klieve et al., 2007). The primary role played by *R. bromii* in releasing energy from RS to other members of the microbial community, and the drop in RS fermentation when this species is absent from the community, justifies designating it as a 'keystone' species within the human colonic microbiota (Ze et al., 2012; Ze et al., 2013).

An initial investigation into the amylases of the human *R. bromii* strain L2-63 concluded that several starch-degrading enzymes are organized into unique multienzyme complexes that we have termed 'amylosomes' which may explain the exceptional RS-degrading activity of this strain (Ze et al., 2015). Amylosome complexes are assembled via interactions between dockerin and cohesin modules present in enzymes and structural proteins (scaffoldins) (Ze et al., 2015) in a manner suggested by the organization of cellulosome complexes implicated in bacterial degradation of lignocellulose (Bayer et al., 2008). Highly elaborate forms of cellulosome organization are found in the related rumen cellulolytic species *Ruminococcus flavefaciens* (Ding et al., 2001; Rincon et al., 2010) and in the human cellulolytic species *Ruminococcus champanellensis* (Ben David et al., 2015; Morais et al., 2016). Amylosome organization has been demonstrated so far only in *R. bromii* L2-63 however and its significance and occurrence across other strains of *R. bromii* is unknown. Indeed, with only a single genome available for analysis until now there has been no information on genetic variation and limited opportunity to probe the microbial ecology of this species. Here we examine the genomic and phenotypic characteristics of four human *R. bromii* strains of diverse origin and of one rumen *R. bromii* strain. This investigation reveals remarkable new insights into the conservation of systems involved in substrate utilization and degradation and into survival mechanisms in this important, but little known, gut symbiont.

Results

Comparative genomics of *R. bromii* strains

Draft genomes were obtained here for three human *R bromii* strains L2-36, 5AMG, ATCC27255 and for the rumen strain YE282. The existing draft genome of the human strain *R. bromii* L2-63 was re-annotated for comparison (Table S1, Figure 1, Experimental procedures). Genome size estimates for the four human strains (2151 to 2400 kb) are slightly lower than for the rumen YE282 strain (2539 kb). The genomic sequences of the *R.*

bromii human strains share 95-100% average nucleotide identity with each other and 86% identity with the rumen strain. Genomes of L2-63, L2-36, 5AMG, ATCC27255 and YE282 *R. bromii* strains were compared using BLAST and the atlases were generated using Blast Ring Image generator (BRIG) software (Figure 1A). A pan-genome analysis identified 735 core genes common to all 5 strains, with *R. bromii* YE282 possessing by far the highest number (1561) of unique genes thus establishing its divergence from the human strains (Figure 1B, Figure S1, Supplemental Data file 1).

Conservation of glycoside hydrolases and carbohydrate-binding modules among *R. bromii* strains

The genomes of the four human *R. bromii* strains encode between 24 and 26 glycoside hydrolases (GHs), and the rumen strain YE282 28 GHs (Supplemental Table S2), with no polysaccharide lyases detectable at the known polypeptide sequence level. This represents an unusually limited CAZyme repertoire for a carbohydrate-utilizing gut bacterium (El Kaoutari et al., 2013) with the great majority of CAZymes belonging to family GH13, and to families (GH23, GH24, GH25) that encode lysozymes (Figure 1C) although the GH23 family is also known to encode chitinases. Among the families that are represented, GH13, GH31 and GH77 enzymes are mostly concerned with starch utilization (Figure 1C). 17 GH13 enzymes were found in all four human *R. bromii* genomes and 15 in the rumen strain YE282 which lacked homologues of Amy3 and Amy7 (Table 1). With three exceptions, genes encoding GH13 enzymes were not clustered (Figure 1A). Phylogenetic analysis revealed remarkable conservation of homologous GH13 sequences as shown in Figure 2 for the 9 extracellular GH13 enzymes that carry N-terminal signal peptides. Furthermore CBM26 modules that have been implicated in starch binding (Boraston et al., 2006) are present in the same 7 extracellular GH13 proteins in all five strains (Table 1).

The four human *R. bromii* genomes, but not the rumen YE282 strain, encode multiple copies of a novel family of accessory ('X') module closest in sequence to CBM37 (Xu et al., 2004) (Supplementary Table S2). In contrast to *R. albus* CBM37-like modules (Xu et al.,

2004), none of the *R. bromii* X modules was found in association with GH domains, cohesins or dockerins. Furthermore, bioinformatic analysis of 50 X modules from *R. bromii* strains revealed that they belong to a phylogenetic clade distinct from the CBM37 modules of *R. albus* (Figure S2). Several of these new *R. bromii* X modules are adjacent to domains annotated as invasin/intimin cell adhesion fragments, cell wall hydrolase /autolysins, and a L,D-transpeptidase.

New insights into amylosome structure and conservation of amylosome components

The formation of enzyme complexes (amylosomes) in *R. bromii* L2-63 was previously inferred from the interactions of dockerin modules present in four extracellular GH13 enzymes with cohesin modules found in four scaffoldin proteins (Ze et al., 2015). The present analysis identifies a fifth dockerin-containing GH13 enzyme, Amy16, and shows that the same five dockerin-carrying enzymes (Amy4, Amy9, Amy10, Amy12 and Amy16) are encoded by all five *R. bromii* strains (Table 1, Supplementary data file 2). Not only the GH13 sequences, but also the size and modular organization of these five enzymes show remarkable conservation between all five strains. Notably, the Amy4 enzyme carries both a dockerin and cohesin module in all strains, although the YE282 enzyme differs in carrying two cohesins and there is some variation also in the number of CBM26 modules carried. Comparison of the activities of the Amy16, Amy4 and Amy9 enzymes from *R. bromii* L2-63 following expression in *E. coli* showed that they acted as α–(1,4)- amylases with no activity against pullulan, indicating a lack of activity against α-(1,6)-linked glucose residues (Figure S3). In contrast, recombinant Amy10 and Amy12 enzymes act as pullulanases that are more active against pullulan than against glycogen and show minimal activity against α-(1,4) linkages.

A fifth scaffoldin, Sca5 was also detected here in the re-annotated L2-63 genome. This comprises two similar modules (51% amino-acids identity) of unknown function, two cohesins and a C-terminal sortase signal and therefore represents a second potential mode

of anchorage, along with Sca2 (Ze et al., 2015), to the bacterial cell surface. Bioinformatic analysis (BLASTP) using L2-63 cohesins and dockerins as queries showed that five similar scaffoldin proteins are present in all five strains (Figure S4). A cohesin from *R. bromii* L2-63 Sca5 was expressed as a fusion protein in *E. coli* to determine its interactions with selected dockerins. The Sca5 cohesin bound strongly to dockerins from Amy4, Amy12 and Amy9, but only weakly to those from Amy10 and Amy16 (Figure 3AB). The Amy16 dockerin did however exhibit strong binding activity for the cohesins from Amy4 (Sca1) and Sca3 (Figure 3AB). We also investigated the interaction of a GFP-tagged dockerin from *R. bromii* L2-63 Amy4 (GFP-doc-13a) with *R. bromii* L2-63 cells. Since dockerin:cohesin interactions require calcium, cells were first treated with EDTA to disrupt protein complexes; subsequent incubation with GFP-doc-13a showed binding in the presence but not in the absence of calcium (Figure 3CDE), thus strongly supporting the hypothesis that binding of Amy4 to the cell surface occurs through dockerin:cohesin interactions. Based on these new data we can now propose a comprehensive scheme for the *R. bromii* L2-63 amylosome that includes Amy16 and a presumed role for the Sca5 scaffoldin whose C-terminal sortase signal indicates that is attached to the cell surface (Figure 4).

We also compared the ability of *R. bromii* strains to utilize RS. The genomes of all five strains indicated a very limited capacity for vitamin synthesis, with only niacin predicted to be synthesised (Magnúsdóttir et al., 2015). For this reason the growth tests were carried out in rumen fluid (M2) medium which was used for the initial strain isolations (Herbeck and Bryant, 1974; Ze et al., 2015). Consistent with the conservation of their amylase systems, the five *R. bromii* strains showed a similar ability to utilize pre-boiled resistant starches. The disappearance of total sugar after 48 h incubation ranged from 77 to 89% for RS2 (mean 82.4 %) and from 71 to 87% for RS3 (mean 78.9%) (Figure 5AB). These figures slightly underestimate the extent of degradation since there was some accumulation of unabsorbed reducing sugar, as noted previously for *R. bromii* L2-63 (Ze et al., 2012).

Capacity for sugar utilization by *Ruminococcus bromii* strains

The five *R. bromii* genomes encode limited sets of carbohydrate transporters. An ABC transport system predicted to take up maltose and maltodextrins, the major products of starch hydrolysis, is encoded by genes upstream of the Amy5 amylase (Figure S5). Separately, linked *malP* and *malQ* genes encode maltodextrin phosphorylase and a GH77 glucanotransferase, suggesting that maltose is metabolized via a phosphorylase/ glucanotransferase cycle similar to that in *E. coli* (Boos and Shuman, 1998). A fructose-specific phosphotransferase gene cluster (Figure S5) and a galactose utilisation gene cluster including a Na^+ galactoside symporter (melB) are also present in the four human *R. bromii* genomes (Figure 5C). The ability to grow with galactose as sole energy source, not previously identified in this species, was confirmed experimentally, while only strain L2-36 grew with glucose as sole energy source (Figure 5D). In contrast, the genome of the rumen strain YE282 lacks a close homologue of *melB* and carries an incomplete fructose utilization cluster. YE282 grew well on maltose and starch but failed to utilize any of the monosaccharides tested suggesting that it must rely solely on the uptake of oligosaccharides (Figure 5D).

Sporulation, spore survival and germination

The *R. bromii* L2-63 genome encodes 73 genes with inferred functions in sporulation and spore germination. These include the key regulator Spo0A, along with sporulation sigma factors (σH, σF, σE, σG, and σK) and their regulators (the spoIIAA, spoIIGA and spoIIIA operons (Hutchison et al., 2014)) and regulatory proteins (soj and spo0J families) involved in chromosome partitioning (Quisel and Grossman, 2000) (Figure 6A, Supplemental Data file 2). These 73 genes were also detected in *R. bromii* L2-36 and their products were identical in sequence. *R. bromii* 5AMG, ATCC27255 and YE282 lack 3, 8 and 6 of these genes and the corresponding gene products share 98-100%, 81-100% and 30-98% amino acid identity, respectively, with L2-63 homologues (Supplemental Data file 2). The majority of core sporulation genes described recently in *Clostridium difficile* strain 630 and *Bacillus subtilis* strain 168 (Browne et al., 2016) are present in the five *R. bromii* genomes.

We tested the sporulation capabilities of *R. bromii* L2-63 strain by subjecting cultures pre-grown for 72h in M2S liquid medium to heat inactivation (80°C for 20 min) and exposure to air. Restoration of anaerobic conditions resulted in good growth (OD_{650} >0.6) after 24 h incubation in liquid M2S medium containing 0.1% sodium taurocholate. This indicated germination of functional spores that had survived these conditions and 16S rRNA gene sequencing confirmed the identity of the re-grown cultures as *R. bromii*. *R. bromii* endospores were detected by TEM in images obtained after heat treatment of 72 h and 30 day old cultures. *R. bromii* cells showed diffuse cell surface features after 72 h growth in liquid M2S medium (Figure 6B), whereas endospore-like structures were detected in cells grown for 30 days (Figure 6CD) along with release of endospores from *R. bromii* cells (Figure 6EF).

Indications of horizontal gene transfer

An 8-gene vancomycin resistance cluster (Depardieu et al., 2003) was detected in *R. bromii* L2-63 and L2-36 but not in 5AMG, ATCC27255 or YE282 (Figure S6AB). All four human *R. bromii* strains were however fully inhibited by vancomycin at 1 μg/ml, while growth on 0.4 μg/ml vancomycin was similar for all five strains, showing that resistance was not expressed (Figure S6C). The regions flanking the VanG cluster contain genes (XerC site-specific recombinase, TraC, VirB4 and VirD4 Type IV secretion) that suggest acquisition through conjugative-transposon-like elements. Unambiguous CRISPR arrays were found in three of the 5 *R. bromii* strains (ATCC27255, 5AMG and YE282) and possible CRISPR structures in all 5 strains. CRISPR structures are also prevalent in related *Ruminococcus* species (Berg Miller et al., 2012; Wegmann et al., 2014).

Discussion

The organization of carbohydrate-degrading enzyme systems and transporters in polysaccharide-utilizing micro-organisms appears fundamental to their niche specificity and to their competitive ability within the microbial community (Flint et al., 2008). In *R. bromii* the

unique organization of starch-degrading enzymes into extracellular amylosome complexes is proposed to underpin the exceptional ability to degrade dietary resistant starches (Ze et al., 2012; Ze et al., 2013). We have now shown that *R. bromii* L2-63 encodes five dockerin-carrying amylases and five cohesin-carrying scaffoldins (one of which, Amy4, is also an amylase) that participate in these complexes. Two scaffoldins (Sca2, Sca5) carry C terminal sortase signals likely to mediate attachment to the cell surface and a gfp-tagged dockerin from Amy4 (that recognises cohesins in Sca2 and Sca5) was shown here to bind *R. bromii* L2-63 cells in a calcium-dependent manner. Our evidence suggests the formation of alternative cell surface and cell-free scaffoldin-based amylase complexes (Figure 4).

We find that the modular organization of the five dockerin-carrying amylases and of the five scaffoldin proteins is very highly conserved across the five available *R. bromii* strains. These include human strains from the USA (ATCC27255), Ecuador (5AMG) and the UK (L2-63, L2-36), but also a rumen strain from Australia (YE282). While the rumen strain shows the most divergence, its amylosome components nevertheless share the main features found in their homologues in the human strains. This degree of similarity points to the likelihood of an early evolutionary origin for such complexes among specialist starch-utilizing bacteria that colonized the animal gut that predates acquisition of these bacteria by humans. Starch present naturally in the diets of hind-gut fermenting animals will be in the raw state, making it more resistant to the action of mammalian amylases than is starch present in cooked foods consumed by modern day humans. In pre-gastric fermenting ruminants, raw starch present in ingested feed is exposed only to salivary amylases before entering the microbial community of the rumen. Thus the evolution of sophisticated microbial systems capable of degrading a wide range of starches including raw starch particles in the animal gut may well explain the high particulate starch-degrading capacity of human *R. bromii* strains. The significance of fructose and galactose utilization by the human strains, although not in the rumen strain, is unknown, but suggests that these sugars might provide alternatives to starch as an energy source in the human colon. Some *Ruminococcus* strains

from the rumen have been shown to possess an unusual transformation mechanism mediated by membrane vesicles that also contain cellulosome complexes (Klieve et al., 2005). In view of this the possibility that horizontal gene transfer might be mediated by amylosome-containing membrane vesicles in *Ruminococcus bromii* deserves investigation.

Until recently, most of the predominant anaerobic gut bacteria found in gut communities were regarded as 'non-sporing'. The detection of large sets of sporulation genes in many of these species (Browne et al., 2016) has however called this assumption into question. We have shown here that all five *R. bromii* strains possess a complete set of genes required for spore formation and spore germination. In the presence of taurocholate as an inducer of spore germination (Browne et al., 2016), *R. bromii* L2-63 was able to recover from 80^{o}C heat treatment and exposure to air, conditions that kill vegetative cells but are resisted through spore formation. Spore formation has profound consequences for gut microbial ecology as it may help to explain the acquisition of this strictly anaerobic species in human infants around the time of weaning. Sporulation must also increase the potential for acquisition of strains from environmental sources, even in adults; thus although *Ruminococcus* spp. are strongly associated with gut communities rather than non-host-associated environments (La Reau et al., 2016) this does not preclude their transmission between hosts via spores of faecal origin. It has been proposed that for human populations in many industrialized countries gut microbiota diversity is decreasing with each generation as more diverse (typically rural) microbial communities are lost (Bokulich et al., 2016). The transmissibility of key anaerobic species between hosts appears crucial in assessing this hypothesis.

As one of the most dominant bacterial species in the human colon, *R. bromii* has the potential to influence host physiology and health via primary and secondary metabolite production (Donia et al., 2014) and via interactions with the immune system. Furthermore since we know that diet composition has a profound impact on the population of this nutritionally specialised species, it is essential to gain a better understanding of its impact on

the host. This has to start with better understanding of the microbial ecology, physiology and genomics of this little studied bacterium and the present work represents an important step towards this.

Experimental Procedures

Bacterial isolates and growth medium

Human *R. bromii* isolates include ATCC2755 (Moore et al., 1972) and two strains (L2-63 and L2-36) isolated from a healthy male UK child (Barcenilla, 1999; Ze et al., 2012) with approval from Grampian Research Ethics (project no 00/00133). Strain 5AMG strain was isolated at the Rowett of a faecal sample from a healthy individual from rural Ecuador. The rumen strain YE282 was isolated from a steer fed on high grain diet from Australia (Klieve et al., 2007) Strains were routinely maintained in M2GSC medium containing 30% clarified bovine rumen fluid and 0.75% agar (Miyazaki et al., 1997) in Hungate tubes (7.5 ml aliquots). M2 medium supplemented with 0.2% carbohydrate (Miyazaki et al., 1997) was used for growth tests based on OD_{650} of triplicate cultures.

Resistant starch utilisation

Hi-maize 958 is a natural RS2 corn starch and Novelose 330 a retrograded RS3 corn starch (Ze et al., 2012) Raw starch solutions (2% wt per vol) were boiled for 10 min before addition to M2 medium (final concentration 0.2%). Total sugar was determined by the phenol sulphuric acid assay (Dubois et al., 1956) and reducing sugar according to (Lever, 1977).

Sequencing, assembly and annotation of *Ruminococcus bromii* draft genomes

Overnight cultures of *R. bromii* human strains were centrifuged at 14,000 X g for 10 minutes and genomic DNA extracted from cell pellets using FastDNA® Spin Kit for Soil (MP Biomedicals, Cambridge, UK). Sequencing of the *R. bromii* L2-63 genome was done at the Wellcome Trust Sanger Institute, Cambridge UK, by the Pathogen Genomic group (http://www.sanger.ac.uk/resources/downloads/bacteria/metahit/) and of the *R. bromii* YE282

genome at the U.S. Department of Energy Joint Genome Institute (DOE JGI, California) as part of the Hungate 1000 project. These sequences were publicly available. The *R. bromii* L2-63 genome was re-annotated using Prokka (Seemann, 2014) which resulted in an increase of 14.2% in identified coding sequences (CDS) to a total of 2111 compared with 1811 reported previously (Ze et al., 2015).

R. bromii strains L2-36, 5AMG and ATCC2755 were newly sequenced at the Earlham Institute (Norwich, UK), using Illumina HiSeq (Illumina, San Diego, CA, USA) generating paired end reads with read lengths of 50 bp. Libraries were prepared from genomic DNA using the KAPA high throughout Library Prep Kit (Part No: KK8234). Genomic DNA sonicated to an average size of 500 bp, end repaired, tailed, adaptor-ligated, fractionated, purified and enriched. The constructed libraries were normalised and equimolar amounts pooled into one final pool of 7.7 nM (8.1 nM for strain 5AMG) using elution buffer (Qiagen). The library pools were then diluted, transferred into a 200 µL strip tube, spiked with 1% PhiX Control and placed on ice before loading onto the Illumina cBot. Template hybridization and first extension was carried out on the cBot utilising the TruSeq Rapid PE Cluster Kit v1 or HiSeq Rapid PE Cluster Kit v2 prior to the flowcell being transferred onto the HiSeq2500 for the remainder of the clustering process. The sequencing chemistry used was TruSeq Rapid SBS Kit v1 or HiSeq Rapid SBS Kit v2 using HiSeq Control Software 2.2.58 and RTA 1.18.64. The library pool was run in a single lane for 50 cycles of each paired end read. Reads in bcl format were de-multiplexed based on the 6 bp Illumina index by CASAVA 1.8, allowing for a one base-pair mismatch per library, and converted to FASTQ format by bcl2fastq.

The raw data reads were aligned against the *R. bromii* L2-63 genome using the mapping tool Bowtie2 (version 2.2.6). De-novo assembly of the three *R. bromii* strains used ABYSS (version 1.9.0). Assemblies were annotated using Prokka (version1.7.2)(Seemann, 2014)

and employing Pfam (Punta et al., 2012), Prosite (Sigrist et al., 2010) and RNAmmer (Lagesen et al., 2007) to identify protein families and functional protein sites and rRNA.

Data availability

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers NPHY00000000 (*R. bromii* L2-36 strain), NNSR00000000 (*R. bromii* ATCC27255 strain) and NNBY00000000 (*R. bromii* 5AMG strain). The version described in this paper is version NPHY01000000, NNSR01000000 and NNBY01000000 respectively.

Genome analysis

Blast Ring Image generator (BRIG) software was used for genome comparisons (Alikhan et al., 2011). Pangenome analysis used BPGA software Version 1.3 (Chaudhari et al., 2016). Orthologous genes (OGs) present in all five *R. bromii* genomes were defined as core genes, sequences with orthologs present in 2-4 strains were considered as variable or accessory genes and sequences that were found only in one strain were considered unique genes. Analysis of the Carbohydrate-active enzymes (CAZymes) relied on the CAZY website (http://www.cazy.org/) supplemented by Hidden Markov models (Sheridan et al, 2016). Cohesin and dockerin sequences were predicted using the BLASTP and tBLASTn algorithm (Altschul et al., 1997) utilising known dockerin and cohesin query sequences. Hits of E-value $<10^{-4}$ were individually examined. Signal Peptide sequences were predicted using the SIGNALP server (http://www.cbs.dtu.dk/services/SignalP/). ClustalW was used for multiple sequence alignments and phylogenetic trees constructed by the maximum-likelihood method, using MEGA7.0 software (Kumar et al., 2016). Promoter sequences were identified using BPROM (http://www.softberry.com) and Rho-independent transcription terminators predicted using ARNold (http://rna.igmors.u-psud.fr/toolbox/arnold/index.php). Antibiotic resistance genes were analyzed through the Comprehensive Antibiotic Resistance Database (CARD) (version 1.0.3) using BLASTP (McArthur et al, 2013). Clustered Regularly

Interspaced Short Palindromic Repeats (CRISPR) arrays were detected by CRISPRFinder web tool (Grissa et al., 2007).

Expression of cohesins and dockerins in *E. coli*

Cloning of the xylanase fused-dockerin from Amy16, GFP fused-dockerin from Amy4 (GFP-doc-13a) and CBM-fused Sca5 cohesin, expression in *E.coli* BL21 (DE3) cells and purification of the recombinant proteins were performed following previously published protocol (Barak et al., 2005). Primers used in this study are listed in Supplementary Table S3.

Expression of *R. bromii* amylases in *E. coli*

Full length Amy4, Amy9 and Amy16 sequences were amplified from genomic DNA using the primers listed in Table S4 and cloned into the vectors indicated. Following transformation into *E. coli* BL21 (DE3) and over-expression (Ben David et al., 2015), 6His-tagged products were purified and their activities against alpha-glucan substrates (Supplementary Figures S3AB) assayed by reducing sugar release (Ze et al., 2015). The Amy10 and Amy12 sequences were amplified from genomic DNA to express the soluble portions of the proteins minus the signal peptides or dockerin sequences. These gene fragments were cloned into the LIC vector pETite-Chis (Lucigen, Madison, WI) to include a 6His tag at the C-terminus as well as a TEV cleavage site. The proteins were expressed in Rosetta (DE3) pLysS cells, purified via Ni^{2+}-affinity chromatography, and the 6His tag removed via TEV using methodology previously described (Cockburn et al., 2015). Activity on 0.3% or 0.075% pullulan, and 0.3% glycogen was measured in triplicate (Figure S3C). For activity assays with polysaccharide substrates the production of free reducing ends was monitored using the bicinchoninic acid (BCA) method (Waffenschmidt and Jaenicke, 1987; Cockburn et al., 2015). Assay substrates were: potato starch (Sigma S2004); corn starch (Sigma S4180); glycogen (Sigma G8751); Novelose 330 (National Starch & Chemical Ltd); pullulan (Megazyme).

Binding assays

Cohesin-dockerin binding was measured by ELISA (Barak et al., 2005). The interaction between the GFP-tagged Amy4 dockerin (GFP-doc-13a) protein and *R. bromii* cells was assessed using 24 h cultures of *R. bromii* L2-63 grown in M2S medium (OD_{650}=0.6). After centrifugation the cell bacterial pellet was washed with sterile PBS and divided into three separate tubes. In the first tube, the pellet was re-suspended in PBS, treated with 10mM EDTA and washed with PBS three times. Cells were then treated with GFP-doc-13a protein (40μl) along with 10mM calcium and incubated for 1 h at room temperature on a rotating platform. The cells were then washed with PBS three times to eliminate any unbound protein, centrifuged at 5000 rpm for 10 min, re-suspended in PBS, smeared on microscope slides, air-dried, overlaid with Vectashield anti-fade mounting Medium (Vector Laboratories Ltd, Peterborough, UK) and examined with a Zeiss Axio Observer Z1 inverted fluorescent microscope (Carl Zeiss Ltd, Cambridge, UK) using a Zeiss 10 FITC filter. The *R. bromii* pellet from the second tube was treated exactly as the first tube but without the addition of calcium. The third tube was a control where the bacterial pellet was simply washed in PBS.

Spore germination assay

R. bromii L2-63 cultures grown for 72 h were heated to 80°C for 20 min to kill vegetative cells. Subsequently, the cells were exposed to air, pelleted by centrifugation, washed in sterile PBS and used to inoculate fresh M2S medium containing 0.1% sodium taurocholate. Inoculated tubes were incubated anaerobically at 37°C and OD_{650} checked after 24 h. PCR amplification and sequencing of the full-length 16S rRNA gene was used to confirm species identity.Sporulation images of *R. bromii* strains cultured for 72 h and 30 days were captured using transmission electron microscopy (TEM) (Lawley et al., 2009).

Acknowledgements

Support is acknowledged from BBSRC grant no. BB/L009951/1 and the Scottish Government Food, Land and People program. We thank Hilary Browne (Wellcome Trust Sanger Institute, UK) for re-analysis of the *R. bromii* L2-63 genome and Gillian Donachie for technical help. E.A.B. acknowledges grants from the Israel Science Foundation (ISF) (1349/13) and the United States-Israel Binational Science Foundation (BSF). E.A.B. holds the Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry.

Author contributions: HJF, NJ, WK devised the study. IM, JLG, SM performed the experiments. IM, POS, SM, JLG, HJF, EAB analysed the data and prepared figures and Tables. SHD, PL, AK, DO, WK, AWW, EC, PC, CS, NK, RK, DC provided critical resources and support. HJF, IM, EAB, SM wrote the paper. All authors read and approved the final manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L., and Beatson, S.A. (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**: 402.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Barak, Y., Handelsman, T., Nakar, D., Mechaly, A., Lamed, R., Shoham, Y., and Bayer, E.A. (2005) Matching fusion protein systems for affinity analysis of two interacting families of proteins: the cohesin-dockerin interaction. *J Mol Recognit* **18**: 491-501.

Barcenilla, A. (1999) Diversity of the butyrate-producing microflora of the human gut,. PhD thesis. Robert Gordon University, Aberdeen, UK

Bayer, E.A., Lamed, R., White, B.A., and Flint, H.J. (2008) From cellulosomes to cellulosomics. *The Chem Rec* **8**: 364-377.

Ben David, Y., Dassa, B., Borovok, I., Lamed, R., Koropatkin, N.M., Martens, E.C. et al. (2015) Ruminococcal cellulosome systems from rumen to human. *Environ Microbiol* **17**: 3407-3426.

Berg Miller, M.E., Yeoman, C.J., Chia, N., Tringe, S.G., Angly, F.E., Edwards, R.A. et al. (2012) Phage–bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environ Microbiol* **14**: 207-227.

Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H. et al. (2016) Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* **8**: 343ra382-343ra382.

Boos, W., and Shuman, H. (1998) Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. *Microbiol Mol Biol Rev* **62**: 204-229.

Boraston, A.B., Healey, M., Klassen, J., Ficko-Blean, E., van Bueren, A.L., and Law, V. (2006) A structural and functional analysis of α-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *J Biol Chem* **281**: 587-598.

Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D. et al. (2016) Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**: 543-546.

Chaudhari, N.M., Gupta, V.K., and Dutta, C. (2016) BPGA- an ultra-fast pan-genome analysis pipeline. *Sci Rep* **6**: 24373.

Cockburn, D.W., Orlovsky, N.I., Foley, M.H., Kwiatkowski, K.J., Bahr, C.M., Maynard, M. et al. (2015) Molecular details of a starch utilization pathway in the human gut symbiont *Eubacterium rectale*. *Mol Microbiol* **95**: 209-230.

Depardieu, F., Bonora, M.G., Reynolds, P.E., and Courvalin, P. (2003) The vanG glycopeptide resistance operon from *Enterococcus faecalis* revisited. *Mol Microbiol* **50**: 931-948.

Ding, S.-Y., Rincon, M.T., Lamed, R., Martin, J.C., McCrae, S.I., Aurilia, V. et al. (2001) Cellulosomal Scaffoldin-Like Proteins from *Ruminococcus flavefaciens. J Bacteriol* **183**: 1945-1953.

Donia, M.S., Cimermancic, P., Schulze, C.J., Brown, L.C.W., Martin, J., Mitreva, M. et al. (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**: 1402-1414.

Dubois, M., Gilles, K., Hamilton, J., Rebers, P., and Smith, F. (1956) Colorimetric Method for Determination of Sugars and Related Substances. *Anal Chem* **28**: 35-356.

El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., and Henrissat, B. (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* **11**: 497-504.

Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., and White, B.A. (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* **6**: 121-131.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35**: W52-57.

Herbeck, J.L., and Bryant, M. (1974) Nutritional features of the intestinal anaerobe *Ruminococcus bromii*. *Applied microbiology* **28**: 1018-1022.

Hutchison, E.A., Miller, D.A., and Angert, E.R. (2014) Sporulation in Bacteria: Beyond the Standard Model. *Microbiol Spectr* **2**. doi: 10.1128/microbiolspec. TBS-0013-2012.

Klieve, A.V., O'Leary, M.N., McMillen, L., and Ouwerkerk, D. (2007) *Ruminococcus bromii*, identification and isolation as a dominant community member in the rumen of cattle fed a barley diet. *J Appl Microbiol* **103**: 2065-2073.

Klieve, A.V., Yokoyama, M.T., Forster, R.J., Ouwerkerk, D., Bain, P.A., and Mawhinney, E.L. (2005) Naturally occurring DNA transfer system associated with membrane vesicles in cellulolytic *Ruminococcus* spp. of ruminal origin. *Appl Environ Microbiol* **71**: 4248-4253.

Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**: 1870-1874.

La Reau, A.J., Meier-Kolthoff, J.P., and Suen, G. (2016) Sequence-based analysis of the genus *Ruminococcus* resolves its phylogeny and reveals strong host association. *Microbial Genomics* **2**.

Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., and Ussery, D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**: 3100-3108.

Lawley, T.D., Croucher, N.J., Yu, L., Clare, S., Sebaihia, M., Goulding, D. et al. (2009) Proteomic and genomic characterization of highly infectious *Clostridium difficile* 630 spores. *J Bacteriol* **191**: 5377-5386.

Lever, M. (1977) Carbohydrate determination with 4-hydroxybenzoic acid hydrazide (PAHBAH): effect of bismuth on the reaction. *Anal Biochem* **81**: 21-27.

Magnúsdóttir, S., Ravcheev, D., de Crécy-Lagard, V., and Thiele, I. (2015) Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front Genet* **6**: 148.

Miyazaki, K., Martin, J.C., Marinsek-Logar, R., and Flint, H.J. (1997) Degradation and utilization of xylans by the rumen anaerobe Prevotella bryantii (formerly *P. ruminicola* subsp. brevis) B(1)4. *Anaerobe* **3**: 373-381.

Moore, W., Cato, E., and Holdeman, L. (1972) *Ruminococcus bromii* sp. n. and Emendation of the Description of Ruminococcus Sijpestein *Int J Syst Evol Microbiol* **22**: 78-80.

Morais, S., Ben David, Y., Bensoussan, L., Duncan, S.H., Koropatkin, N.M., Martens, E.C. et al. (2016) Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, *Ruminococcus champanellensis*, reveals a fine-tuned system for cohesin-dockerin recognition. *Environ Microbiol* **18**: 542-556.

Nugent, A.P. (2005) Health properties of resistant starch. *Nutrition Bulletin* **30**: 27-54.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C. et al. (2012) The Pfam protein families database. *Nucleic Acids Res* **40**: D290-301.

Quisel, J.D., and Grossman, A.D. (2000) Control of sporulation gene expression in *Bacillus subtilis* by the chromosome partitioning proteins Soj (ParA) and Spo0J (ParB). *J Bacteriol* **182**: 3446-3451.

Rincon, M.T., Dassa, B., Flint, H.J., Travis, A.J., Jindou, S., Borovok, I. et al. (2010) Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefacien*s FD-1. *PLoS One* **5**: e12476.

Salonen, A., Lahti, L., Salojarvi, J., Holtrop, G., Korpela, K., Duncan, S.H. et al. (2014) Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *ISME J* **8**: 2218-2230.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.

Sigrist, C.J., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A., and Hulo, N. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* **38**: D161-166.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L.G., Gratadoux, J.-J. et al. (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *PNAS* **105**: 16731-16736.

Suau, A., Rochet, V., Sghir, A., Gramet, G., Brewaeys, S., Sutren, M. et al. (2001) *Fusobacterium prausnitzii* and related species represent a dominant group within the human fecal flora. *Syst Appl Microbiol* **24**: 139-145.

Waffenschmidt, S., and Jaenicke, L. (1987) Assay of reducing sugars in the nanomole range with 2, 2′-bicinchoninate. *Anal Biochem* **165**: 337-340.

Walker, A.W., Ince, J., Duncan, S.H., Webster, L.M., Holtrop, G., Ze, X. et al. (2011) Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J* **5**: 220-230.

Wegmann, U., Louis, P., Goesmann, A., Henrissat, B., Duncan, S.H., and Flint, H.J. (2014) Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota (‘*Ruminococcus bicirculans*’) reveals two chromosomes and a selective capacity to utilize plant glucans. *Environ Microbiol* **16**: 2879-2890.

Xu, Q., Morrison, M., Nelson, K.E., Bayer, E.A., Atamna, N., and Lamed, R. (2004) A novel family of carbohydrate-binding modules identified with *Ruminococcus albus* proteins. *FEBS letters* **566**: 11-16.

Ze, X., Duncan, S.H., Louis, P., and Flint, H.J. (2012) *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J* **6**: 1535-1543.

Ze, X., Le Mougen, F., Duncan, S.H., Louis, P., and Flint, H.J. (2013) Some are more equal than others: the role of "keystone" species in the degradation of recalcitrant substrates. *Gut Microbes* **4**: 236-240.

Ze, X., Ben David, Y., Laverde-Gomez, J.A., Dassa, B., Sheridan, P.O., Duncan, S.H. et al. (2015) Unique organization of extracellular amylases into Amylosomes in the resistant starch-utilizing human colonic Firmicutes bacterium *Ruminococcus bromii*. *MBio* **6**: e01058-01015.

Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T. et al. (2016) Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**: 565-569.

Figure legends

Figure 1. Comparison of five *R. bromii* genomes. (A) Pan-genome analysis of core, accessory and unique genes among five *R. bromii* strains. A total of 735 core orthologous groups (OGs), 1456 variable OGs and 2240 unique genes were detected in the *R. bromii* pan-genome. Overall, the *R. bromii* YE282 strain genome has the highest number of unique genes (1561) compared to the human *R. bromii* genomes suggesting that these may be associated with colonization of the rumen. (B) Genomic comparison of four *R. bromii* genomes to reference strain ATCC27255. Diagram represents BLASTn results of each genome against ATCC2755 strain with results rendered using the BRIG program. The inner circle represents the reference genome *R. bromii* ATCC27255. Each genome is colour coded as indicated by the legend. Relative shading density (from darker to lighter) within each circle represents relative levels of nucleotide homology. White regions indicate regions with no identity to the reference. The location of genes encoding GH13 enzymes (Amy1-17) is also indicated. (C) Representation (number per genome) of glycoside hydrolase families encoded by *R. bromii* genomes.

Figure 2. Phylogenetic tree based on the amino acid sequences of GH13 catalytic domains of the 9 extracellular GH13 enzymes (that carry signal peptides) from 5 *R. bromii* strains. The sequences of the GH13 genes shown here from the five *R. bromii* strains fall into 9 significant clusters, with sequences from the four human *R. bromii* strains clustering more closely with each other than the YE282 rumen strain. YE282 sequences are identified here by their numerical identifiers in the JGI database, whereas genes from the human strains are prefixed by the strain designation. Bootstrap values, expressed as a percentage of 1000 replications, are given at the branching nodes. This tree is unrooted and reconstructed using the maximum-likelihood method. The scale bar refers to the number of amino acid differences per position. Similar clustering was also observed for the remaining GH13 enzymes listed in Table 1 that are common to all five strains (not shown).

Figure 3. **Cohesin-dockerin binding measured by ELISA**. (A) A microtitre plate was coated with XyndocAmy16. Positive interactions of the Amy16 dockerin were observed with cohesins 1 and 3. (B) ELISA plates were coated with CBM-Coh1 from Sca5. Positive interactions of the cohesin 1 from Sca5 were observed with Xyndocs 14720, 16032, Amy4, Amy12 and Amy9, low interaction was also observed with Xyndoc Amy10. (Please note that to allow comparison the CBL numbering for Xyndocs given here is consistent with that given previously [6] and does not refer to the L2-63 genome re-annotation). Error bars indicate SD from the mean of duplicate samples from one experiment. (C) Recombinant gfp-Amy4doc protein incubated with EDTA-pretreated, washed *R. bromii* L2-63 cells in the presence of Ca^{2+}. (D) as (C) but in the absence of Ca^{2+}. (E) *R. bromii* L2-63 cells in the absence of gfp-amy4doc.

Figure 4. Updated model for cell-bound and cell-free amylosome complexes in *R. bromii* L2-63. The Amy4 and Amy9 enzymes are likely to bind to the cell surface via the Sca2 scaffoldin protein. Amy4 has then the potential to self-aggregate through interactions between its own cohesin and dockerin or to integrate Amy9, Amy10 or Amy16. Further complexes are likely to form between the Amy4, Amy9, Amy10, Amy12, Amy16 and either the Sca3 or Sca4 proteins to form cell-free amylosomes. In addition, Amy4, Amy9, Amy10, Amy12 can be integrated to the first cohesin of Sca5 and attached to the cell-surface. The binding specificities of cohesin modules of the second Sca5 cohesin (shown in light gray) is yet to be determined.

Figure 5. Carbohydrate utilization in *R. bromii* strains. Total sugar utilized and the concentration of free soluble reducing sugar (glucose equivalents) are shown after 48 h incubation in M2 medium containing (A) RS2 (High-maize 958) or (B) RS3 starch (Novelose 330). Data plotted in graphs are the mean ± SD OD readings (OD_{490} for total sugar assay and OD_{415} for reducing sugar assay) of three biological replicates and three technical replicates for each time-point studied for each strain (C) Organization of the galactose operon in the human *R. bromii* strains (P indicates likely promoters and the hairpin indicates

the likely transcriptional terminator). (D) Growth of the five *R. bromii* strains with soluble potato starch, maltose, fructose, galactose, lactose, glucose and arabinose (0.2% w/v) as sole sources of carbon. Data plotted in graphs are the mean OD_{650} readings of three replicates of each strain grown on different sugar substrates. Since the SD values were very small in all cases they were not plotted on to the graph.

Figure 6. Sporulation gene signature and TEM of *R. bromii* sporulation. (A) A set of 65 genes with known sporulation function were detected in the genome of *R. bromii* L2-63 strain. These genes corresponded to: initiation of spore formation (Stage I), asymmetrical division into a larger mother cell and a smaller forespore (Stage II), engulfment of the forespore by the mother cell (Stage III), formation of the spore cortex (Stage IV), deposition of the spore coat (Stage V) followed by lysis of the mother cell and release of the endospore (Stage VI). A set of 8 genes related to germination of the endospore were also detected. Spo0A, which is a critical transcription factor to initiate sporulation and the specialized sporulation sigma factors (σH, σF, σE, σG, and σK); small proteins that directs RNA polymerase to specific sites on DNA to initiate gene expression) along with the regulators of these sigma factors (such as spoIIAA, spoIIGA and the spoIIIA operon) were also present. Key regulatory proteins involved in chromosome partitioning, belonging to the soj (parA_1, parA_2, parA_3,) and spo0J (parB_1, parB_2, parB_3, parB_4, parB_5) families were also detected. In the absence of spoJ, soj is known to negatively regulate expression of several sporulation genes by binding to the promoter regions and inhibiting transcription indicating a tightly regulated energy-intensive process for survival. Comparison with core sporulation and germination related genes from *Clostridium difficile* strain 630 and *Bacillus subtilis* strain 168 described recently by Browne *et al* (Browne et al., 2016) showed that all the key sporulation genes were present. (B) TEM image of a *R. bromii* L2-63 strain after 72 h growth on M2S medium. (C and D) Endospores were visible from *R. bromii* cells (E) Release of endospores from *R. bromii* after lysis of the mother cell. (F) Endospores released into the surrounding medium.

Supporting information:

Figure S1. Pan genome analysis of the five *R. bromii* genomes. A) Phylogenetic analysis using BPGA software based on concatenated core gene alignments. B) COG distribution of core, accessory and unique genes of *R. bromii* strains as deduced by the pan-genome analysis.

Figure S2. Phylogenetic tree of CBM37 modules from *R. albus* and 'X' modules from *R. bromii.*

Figure S3. Activities of recombinant enzymes expressed in *E. coli.*

Figure S4. Phylogenetic tree comparing the 30 cohesin modules from *R. bromii* L2-63, L2-36, 5AMG, ATCC27255 and YE282 strains.

Figure S5. Carbohydrate transport/utilisation gene clusters in *R. bromii* genomes. Organisation of (A) maltose and fructose transport systems and (B) glycogen biosynthesis gene cluster for the human *Ruminococcus bromii* L2-63, L2-36, ATCC27255 and 5AMG strains.

Figures S6. Vancomycin gene cluster in *R. bromii* L2-63. A) B) This VanG cluster is present in *R. bromii* L2-63 and L2-36, but not in the other three genomes. However we found no evidence for resistance to vancomycin in the two human strains that possess the VanG cluster compared with the two that lack the cluster (C).

Table S1. *Ruminococcus bromii* genome assembly statistics.

Table S2. Distribution of GH, dockerins, cohesins and CBM modules in human and rumen *Ruminococcus bromii* strains

Table S3. List of primers used for the different constructs produced.

Supplemental Data file 1. List of core, accessory and unique genes from pan-genome

analysis of 5 *R. bromii* strains.

Supplemental Data file 2. List of locus tag numbers for amylase scaffoldin and dockerin genes for *R. bromii* L2-63 strain before and after re-annotation of the genome (Worksheet1), list of sporulation genes detected in *R. bromii* L2-63 strain (Worksheet 2), conservation of L2-63 sporulation genes in 4 other *R.bromii* strains (Worksheet 3), comparison of spore signature genes described by Browne *et al* with *R. bromii* strains (Worksheet 4).

Table 1. Conserved modular organization of *Ruminococcus bromii* gene products that include GH13 glycoside hydrolase domains.

	size (aa) human	size (aa) rumen	Signal peptide	Cohesin domain	Dockerin domain	CBM 26	CBM 48	Presumed function	Amylosome-related domain architecture
Amy 1	805-804	797	+	-	-	+ (1)	-	amylase	SP CBM26 **GH13**
Amy 2	752-751	748	+	-	-	+ (1)	-	amylase	SP CBM26 **GH13**
Amy 3*	630-629	-	-	-	-	-	-	amylase	CBM34 **GH13**
Amy 4	1357-1356	1528	+	+	+	+ (2)	-	amylase	SP **GH13** CBM26 CBM26 Coh Doc
Amy 5	565-551	554	+	-	-	-	-	amylase	SP **GH13**
Amy 6	512-511	510	-	-	-	-	-	amylase	**GH13**
Amy 7*	428-427	-	-	-	-	-		amylase	**GH13**
Amy 8	565-552	545	-	-	-	-	-	amylase	**GH13** DUF
Amy 9	1064-1056	801	+	-	+	+ (1)§	-	amylase	SP **GH13** CBM26 Doc
Amy 10	1234-1233	1333	+	-	+	+ (2)†	+ (1)	pullulanase	SP CBM48 **GH13** MucBP MucBP CBM26 MucBP Doc CBM26
Amy 11	941-940	961	-	-	-	-	+ (1)	pullulanase	CBM48 **GH13** MucBP MucBP MucBP
Amy 12	1060-1059	1059	+	-	+	+ (1)	+ (1)	pullulanase	SP CBM48 **GH13** MucBP Doc MucBP CBM26
Amy 13	641-639	632	-	-	-	-	+ (1)	glycogen	CBM48 **GH13**
Amy 14	767-762	717	-	-	-	-	+ (1)	glycogen	CBM48 **GH13**
Amy 15	695-694	695	-	-	-	-	+ (1)	glycogen	CBM48 **GH13**
Amy 16	877-867	995	+	-	+	+ (2)†	-	amylase	SP **GH13** CBM26 Doc CBM26
Amy 17	569-556	665	+	-	-	+ (1)‡	-	amylase	SP **GH13**

* Amy3 and Amy7 were absent in rumen strain YE282, § absent in YE282, †3 modules present in YE282, ‡only present in YE282.
SP= Signal Peptide, CBM=Carbohydrate binding module, GH= Glycoside hydrolase, DUF= Domain of Unknown Function, MucBP= Mucin Binding Protein

Figure 1

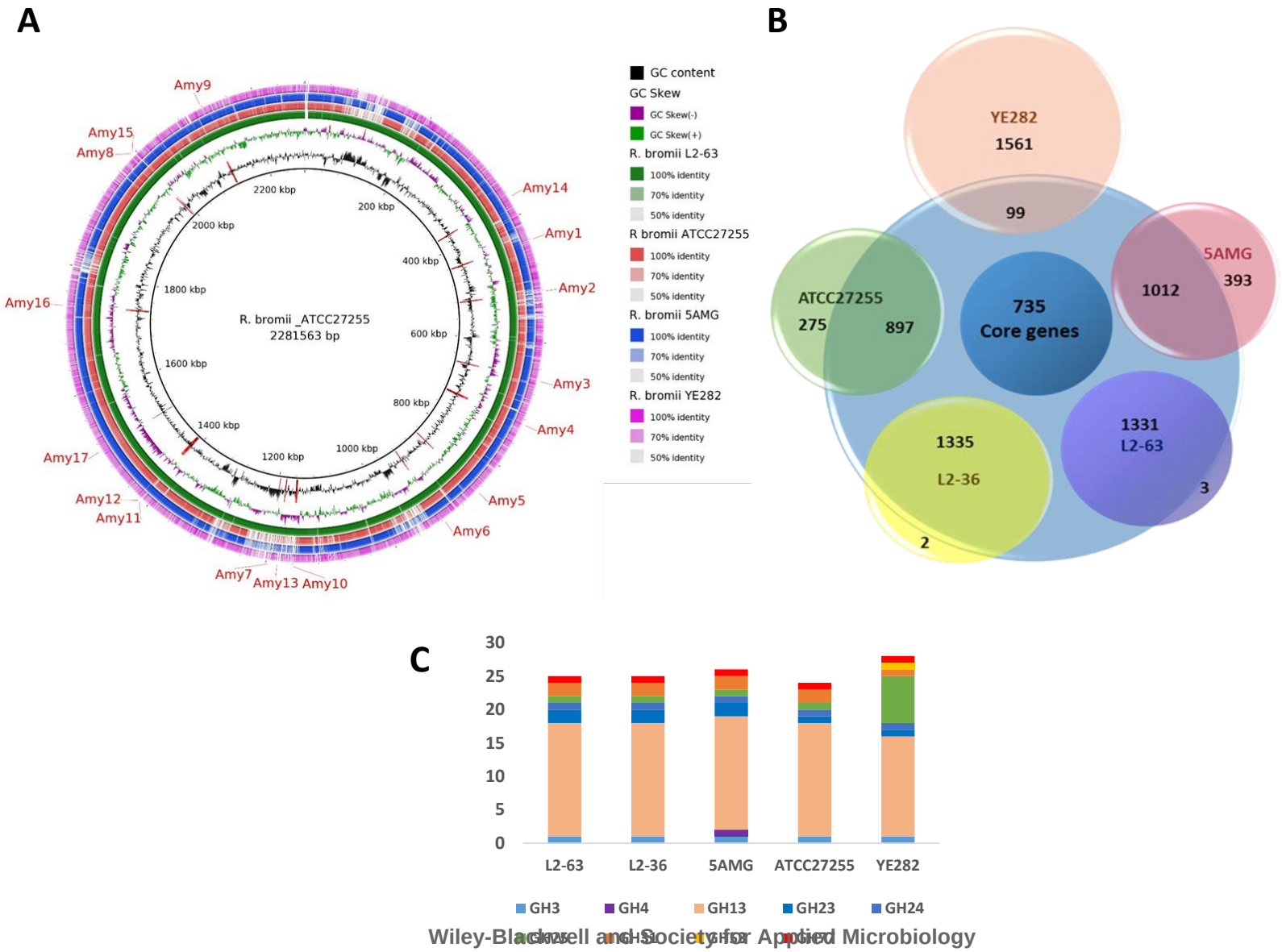


Figure 2

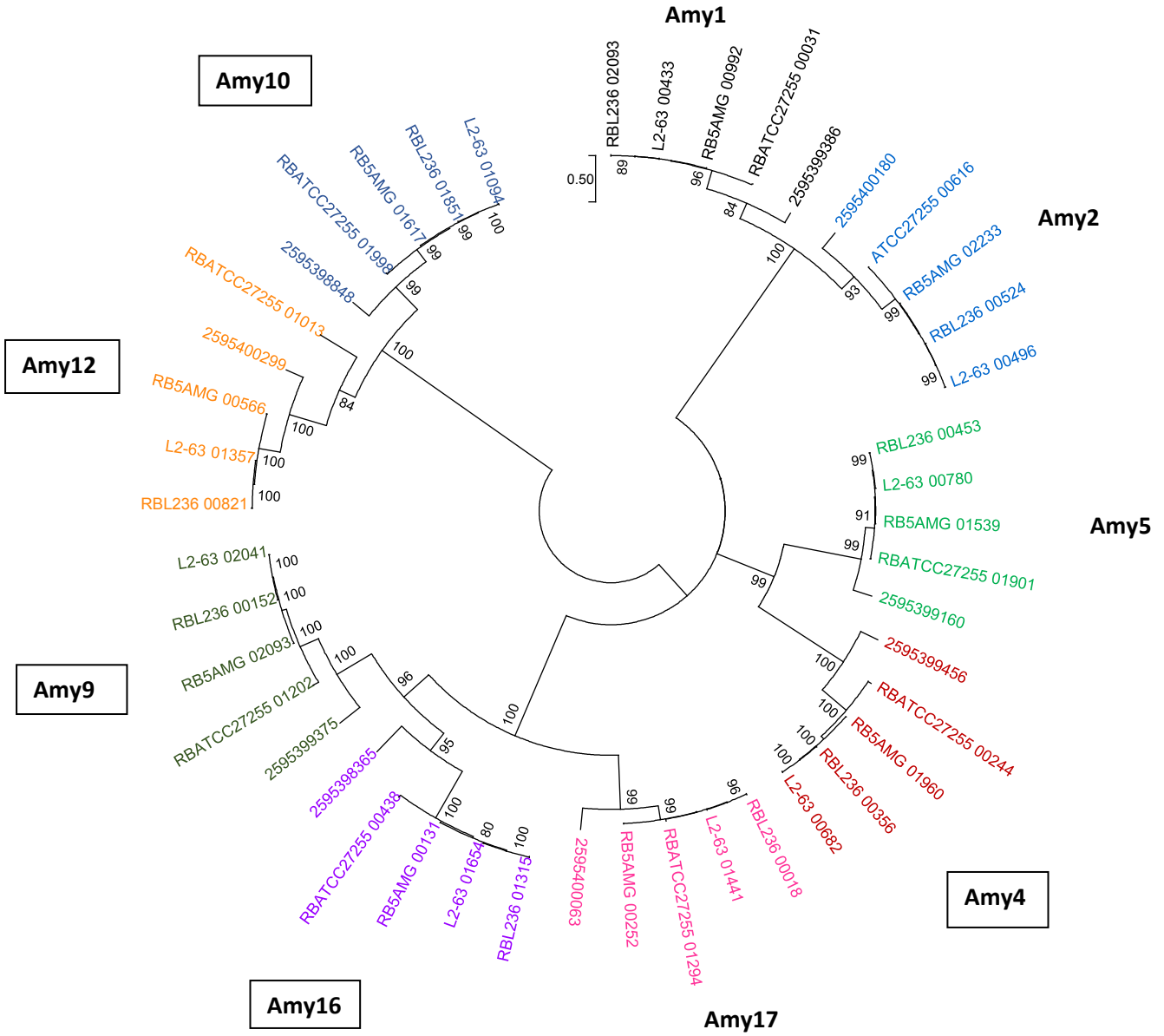


Figure 3

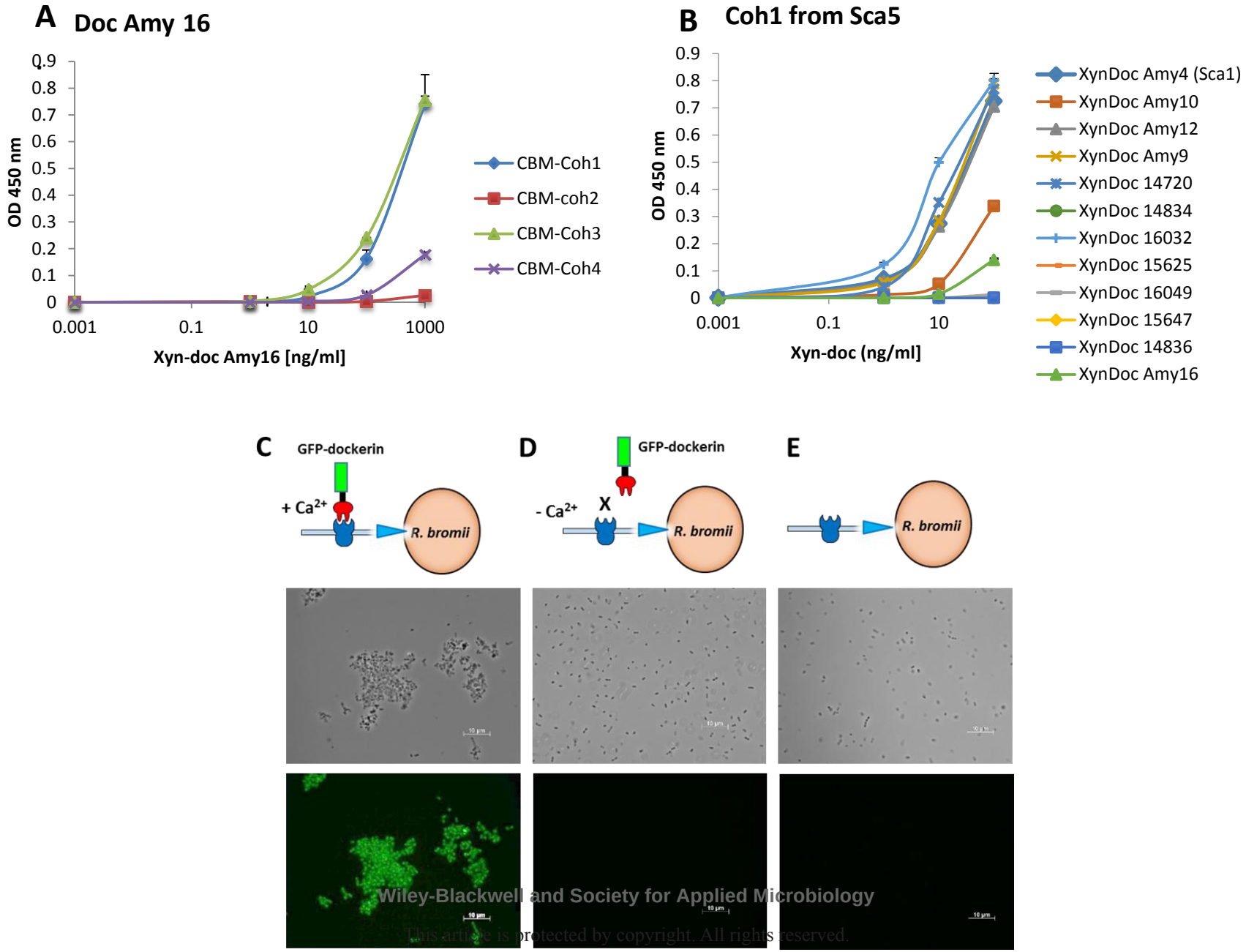


Figure 4

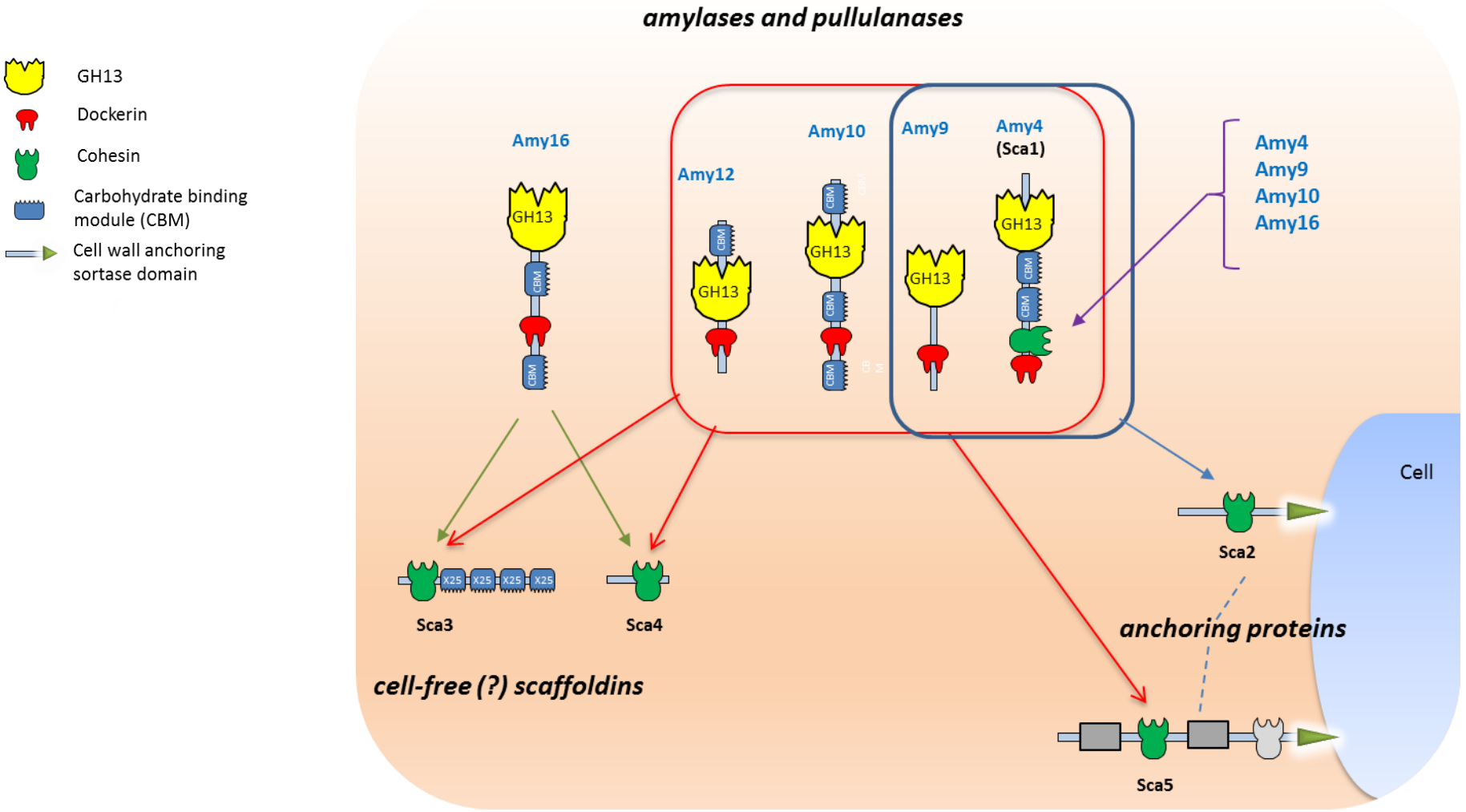


Figure 5

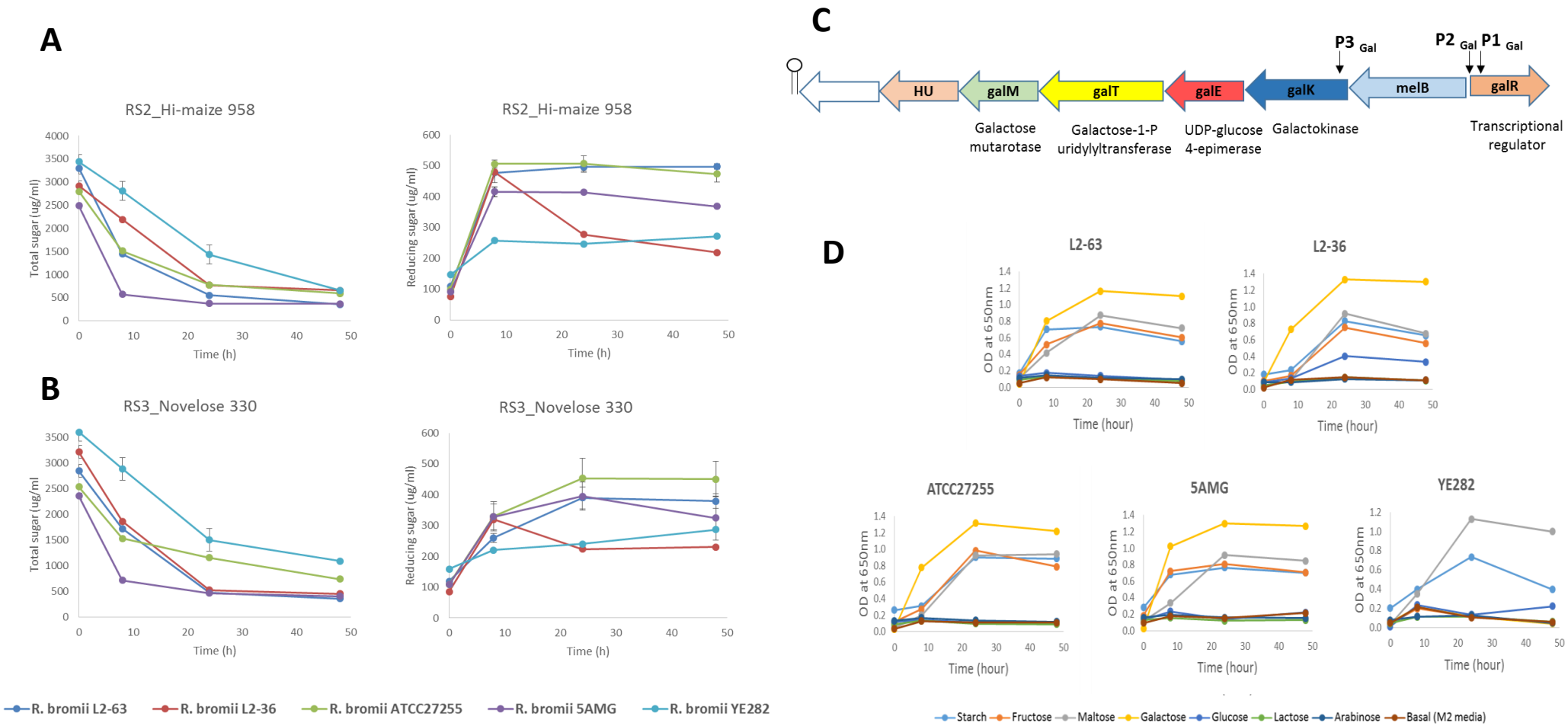
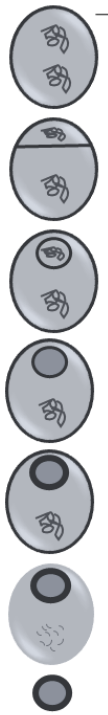


Figure 6

A



Stages of sporulation	Function	Sporulation genes in *R. bromii* L2-63	Sporulation gene signature (Browne et al, 2016, Nature)
Stage I	Initiation of spore formation	sigH, spo0A, spo0J family (parB_1, parB_2, parB_3, parB_4, parB_5), sepF	sigH, Spo0A, parB, sepF,
Stage II	Asymmetric cell division	sigE, sigF, spoIIGA, spoIIAA_1, spoIIAA_2, spoIIAB, spoIIR, spoIIQ, spoIIE, rsbU, spoIIP, L2-63_01395,	sigE, sigF, mcsA, spoIIGA, spoIIR, spoIIp, ykvI
Stage III	Engulfment of forespore by mother cell	spoIIIAB, spoIIIAC, spoIIIAD, spoIIIAE, spoIIIAG, spoIIIAH, spoIIIE, sigG, L2-63_01811	spoIIIAB, spoIIIAC, spoIIIAD, spoIIIAE, spoIIIAG, spoIIIAH, spoIIIAA, ftsQ, sigG, CD630_10450, CD630_20350
Stage IV	Formation of cortex	sigK, spoIVA, spoIVB, YqfD, YlbJ	sigK, spoIVA, spoIVB, spoIVFB, YlbJ, dacF1, dacF, dapG
Stage V	Formation of coat	spoVAE, spoVAC, spoVAD, spoVFA, spoVT, ytgP, ftsI, ftsW_2, ugtp, CotJA, CotJB, SafA	spoVAD, spoVE, spoVFA, cotG, cotCB, sipL, soda, spoVT, CD630_34940, CD630_35690
Stage VI	Mother cell lysis, release of spore	-	-
Germination	Germination of spore	gpr, GerA,GerC sleB, cwlJ, spmA, spmB, L2-63_01843	gpR, spmA, spmB, tepA
Sporulation initiation inhibitor	Regulation	soj_1, soj_2, soj_3, KapD	-
Unknown sporulation genes	Genes not associated with any specific stage	paiA, yebN, xynD, ytfJ, YlmC. YunB, yhhT, YabP, whiA, Yqfc, sspC2, L2-63_00863	asrC, dnaA, PHP domain, bltD, pdp, dnaA, ylxY, rplJ, trxB, ybbH, walR, purB, zupT, CD630_25890, 22730, 05720, 01290, 20480, 00550, 21080, 08810, 24430